Optimal time-frequency bases for EEG signal classification in BCI context

Sebastiano Barbieri, Bruno Torrésani

Optimal Time-Frequency Bases for EEG Signal Classification in the Context of BCI

Sebastiano Barbieri, Bruno Torrésani

Aix-Marseille Université, CNRS, Centrale Marseille, LATP, UMR 7353, 13453 Marseille, France
Sebastiano.Barbieri@cmi.univ-mrs.fr, Bruno.Torresani@univ-amu.fr

Résumé – Nous considérons le problème de classification de signaux multicapteurs, plus particulièrement de signaux EEG dans un contexte BCI, par sélection de caractéristiques temps-fréquence. Les caractéristiques sont recherchées sous la forme de bases de cosinus locales (MDCT) via un algorithme de type "best basis" adapté au contexte de classification. L'algorithme est validé sur des données BCI d'imagination motrice.

Les techniques de décision généralement développées pour le contexte BCI, ou plus généralement en neurosciences, sont souvent basées sur des approches automatiques (réseau de neurones artificiels, SVM,...) ne permettant pas en retour une interprétation simple des caractéristiques. L'approche proposée permet un tel retour sur l'interprétation, car les caractéristiques sont recherchées sous la forme d'atomes temps-fréquence, proches des interprétations classiques en analyse d'EEG, qui font intervenir des bandes de fréquence et des instants spécifiques.

L'algorithme proposé étend l'algorithme *best discriminant basis* de Saito en exploitant des comparaisons deux à deux des signaux des deux bases de données (les deux classes). Les résultats de simulations numériques montrent que la méthode est capable de retrouver des caractéristiques simulées. Les résultats sur données réelles sont tout à fait compétitifs par rapport à l'état de l'art en termes de classification, et plus facilement interprétables.

Abstract – We consider the problem of classifying multi-sensor signals, more precisely EEG signals in the context of Brain Computer Interfaces (BCI), by selection of time-frequency features. The features are determined among local cosine bases (MDCT) by a "best basis" type algorithm adapted to the classification context.

In the BCI domain, or more generally in neuroscience, many classification algorithms are based upon automatic approaches (artificial neural networks, SVM, ...) which do not allow a simple interpretation of the features. The proposed approach allows such interpretation, since the features are determined in the form of time-frequency atoms, similarly to classic analyses of EEG signals which involve specific frequency bands and time intervals.

The proposed algorithm generalizes the *best discriminant basis* algorithm by Saito, employing pairwise comparisons between the signals belonging to two classes of data. Results on artificial data show that the method is able to determine simulated differences between signals. Results on real data are competitive with state of the art classification algorithms and more easily interpretable.

1 Introduction and Problem Statement

Brain computer interfaces (BCI) are devices that allow transforming human intentions and cognitive states (characterized by specific brain signals) into commands. The interpretation of such signals relies on specific features, which are generally estimated during a training stage. In many successful approaches, these extracted features do not allow an easy understanding in terms of the underlying neurophysiological mechanisms.

In this work the classification of a signal is based upon features whose time-frequency localization properties are learnt from a training set. This approach exploits the so called *best basis paradigm* developed by Coifman and Wickerhauser [1, 7], which selects an optimal basis from a *library* of orthonormal time-frequency bases. It also relies on the approach developed by N. Saito in his PhD thesis [5] (see also [6] for an adaptation to the BCI context), and generalizes it by replacing the comparison of average signals by systematic pairwise comparisons of signals belonging to two classes.

The problem at hand is a typical BCI problem, but can be transposed to many different domains. The training set consists of observations $\mathbf{X} = \{(\mathbf{x}_i, \kappa_i)\ , i = 1 \dots N_i\}$, where each $\mathbf{x}_i$ is a realization of a multi-sensor signal $\mathbf{x} = \{x^{(c)} \in \mathbb{R}^N, c = 1 \dots N_c\}$, and $\kappa_i = 0, 1$ is the class label. The working assumption is that the features that allow discriminating between the two classes possess specific but unknown time-frequency signatures. More precisely: they can be characterized by specific localization properties which one seeks to define.

Albeit not pursued in this article, a more ambitious formulation of the problem involves the additional search for topographic information that allows class discrimination, resulting in a simultaneous spatial, spectral, and temporal filter.

2 Discriminant Local Trigonometric Bases

The characterization of discriminant time-frequency localization properties is tackled via the *best basis* approach, applied to local trigonometric (also called MDCT) bases.

2.1 Local Trigonometric Bases

We shall work in the reference framework of the Euclidean space $\mathcal{H} = \mathbb{R}^L$ of finite length signals. In full generality, a local trigonometric basis is obtained through a segmentation of the time axis, using smooth windows, followed by a cosine expansion of the so-obtained segments. More precisely, the integer interval $\mathbb{Z}_L = \{0, \dots L-1\}$ is split into T sub-intervals $I_\tau = [a_\tau, a_{\tau+1}]$ of length $\ell_\tau = a_{\tau+1} - a_\tau$; Given a family of windows ψ_τ, $\tau = 0 \dots T-1$ whose support is essentially concentrated inside the interval I_τ, and satisfying specific compatibility conditions (see [7]), the family of waveforms $\psi_{\tau n}$ defined by

$$\psi_{\tau n}[t] = \sqrt{\frac{2}{\ell_\tau}} \psi_\tau[t] \cos\left[\pi \left(n + \frac{1}{2}\right) \frac{t - a_\tau}{\ell_\tau}\right] \qquad (1)$$

is an orthonormal basis of $\mathbb{R}^L$. In a time-frequency plane, $\psi_{\tau n}$ is essentially localized inside a domain which can be represented by a rectangle $I_\tau \times J_{\tau n}$, where $J_{\tau n}$ is an integer interval in the frequency domain centered on the frequency $\pi(n + 1/2)/\ell_\tau$ and of width π/ℓ_τ. Various choices for the segmentation lead therefore to different (approximate) tilings of the time-frequency plane.

2.2 Library of Bases and Selection of the Optimally Discriminant Basis

The *best basis* algorithm is based on a tree-structured family of splittings of the integer interval $\mathbb{Z}_L$ into subintervals I_{jk} of length $\ell_{jm} = 2^{-j}L$ (where $j = 0, 1, ...J-1$ is the decomposition level), each splitting giving rise to an orthonormal basis of $\mathbb{R}^L$. We denote by ψ_{jkn} the so-obtained waveforms, and denote generically by $\Lambda(\mathcal{B})$ the index set labelling the waveforms ψ_λ of basis $\mathcal{B}$.

The family of the bases generated that way forms a so-called *library* of orthonormal bases, from which one may seek the optimal one with respect to a given criterion. In the considered situation, the criterion is determined by a contrast function $d : u, v \in \mathbb{R}^+ \to d(x, y) \in \mathbb{R}$. Given two signals $x, y \in \mathbb{R}^L$, assumed to be normalized ($\|x\| = \|y\| = 1$), and an orthonormal basis from the library $\mathcal{B} = \{\psi_\lambda, \lambda \in \Lambda(\mathcal{B})\}$, the contrast between x and y given by a basis $\mathcal{B}$ is defined by

$$D_{\mathcal{B}}(x, y) = \sum_\lambda d\left(|\langle x, \psi_\lambda\rangle|^2, |\langle y, \psi_\lambda\rangle|^2\right) \ . \qquad (2)$$

Such additive contrast functions are well adapted to tree-structured libraries, such as the ones of local trigonometric bases. Indeed in [7] the tree structure allows the evaluation, and thus the optimization, of the criterion over all the bases of the library through dynamic programming. In the numerical results displayed here, we have limited our investigations to the case of the symmetrized Kullback-Leibler divergence

$$d(u, v) = u \ln(u/v) + v \ln(v/u) \ , \qquad (3)$$

however several other choices are possible [5].

2.3 Training and Test

Assume we are given N_x (resp. N_y) signals $x^{(k;c)}$ (resp. $y^{(k;c)}$) corresponding to class X (resp. Y). k is the trial label, and c is the channel label. Channels can be either electrodes, or linear combinations of electrodes determined by CSP, LDA or other discriminant dimension reduction methods. The goal is to determine the basis from the library that allows to optimally discriminate between these two classes, using the measured divergence, for example by means of the metric defined in (3). Two strategies are possible: to determine an optimal basis for each channel, or to determine an optimal basis common to all channels. Herein we limit ourselves to the second possibility, which gave better results on the datasets we considered.

During the training phase, the following operations are performed

- Normalization of the signals:
  $\|x^{k;c}\| = 1$ and $\|y^{k;c}\| = 1$;
  computation of the coefficients of the expansions on all the bases ψ_λ from the library (classes X and Y):
  $$\alpha^{k;c}_{jmn} = \langle x^{k;c}, \psi_{jmn}\rangle \ , \quad \beta^{k;c}_{jmn} = \langle y^{k;c}, \psi_{jmn}\rangle \ . \quad (4)$$
  Thanks to the normalization, whatever the basis $\mathcal{B}$, we have
  $$\sum_{\lambda\in\Lambda(\mathcal{B})} |\alpha^{k;c}_\lambda|^2 = 1 \quad \text{and} \quad \sum_{\lambda\in\Lambda(\mathcal{B})} |\beta^{k;c}_\lambda|^2 = 1 \qquad (5)$$
  for all k, c.
- For each signal pair $(x^{k;c}, y^{\ell;c})$ and each basis vector ψ_λ, computation of the divergence
  $$d^{k\ell;c}[\lambda] =$$
  $$|\alpha^{k;c}_\lambda|^2 \log\left(\frac{|\alpha^{k;c}_\lambda|^2}{|\beta^{\ell;c}_\lambda|^2}\right) + |\beta^{k;c}_\lambda|^2 \log\left(\frac{|\beta^{k;c}_\lambda|^2}{|\alpha^{\ell;c}_\lambda|^2}\right) \ . \quad (6)$$
- For each λ, evaluation of the average divergence $\overline{d}[\lambda]$ (average with respect to signals and channels).
- Determination of the best basis
  $$\mathcal{B}^* = \arg\max_{\mathcal{B}} \sum_{\lambda\in\Lambda(\mathcal{B})} \overline{d}[\lambda] \qquad (7)$$
  and selection of the basis vectors that contribute most significantly to the discrimination, in terms of their average divergence.

During the test phase the coefficients with respect to the optimal basis of each signal from the test dataset are computed. In the simplest case, each signal could be assigned to the class to which the divergence is the lowest. For the results presented in this paper a linear kernel support vector machine (SVM) was used as classifier (LIBLINEAR: [11]).

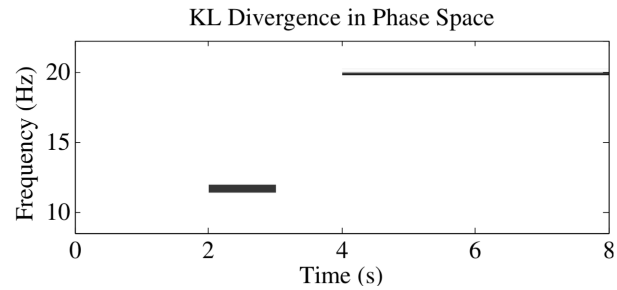

Fig. 1: Magnitude of the symmetrized Kullback-Leibler divergence in phase space for the test on synthetic data.

3 Results

3.1 Data and Pre-Processing

The method has been tested on a motor imagery dataset. Our data consists of 64-channel EEG recordings of 11 subjects. The data was sampled at 2048 Hz and subsequently downsampled to 256 Hz. Each trial started with the user fixating a cross on a screen for 2.5 seconds during which he was allowed to blink. The subject was then presented with a right or left visual cue lasting 1 second. After the appearance of the visual cue the subject had to imagine moving the corresponding ipsilateral hand.

Each subject performed both left and right hand motor imagery 80 times. The order of the trials was random.

The characteristic signals for the considered task are expected to show up in specific frequency bands and locations and correspond to *mu-rhythms* and *beta-rhythms*. Signals have been band-pass filtered to keep the frequency band 4-28 Hz. Both the Common Spatial Pattern [8, 9] and the Common Spatio-Spectral Pattern algorithms [10] were used as spatial filters during pre-processing. As suggested in [9] the output of the 4 most important filters was used for further processing.

3.2 Tests and Results

At first the method is illustrated on simple artificial data. Class X is given by a sample EEG signal of 8 seconds length and class Y is given by the same signal to which a 12 Hz sine wave was added from 2 s to 3 s and a 20 Hz sine wave was added from 4.5 s to 6.5 s. Figure 1 shows that the selected *best basis* contains functions localized in the regions of the time-frequency plane corresponding to the differences between the two signals.

Similar plots are given in Figure 2 for two of the subjects performing motor imagery tasks, illustrating the tiling of the time-frequency plane and the magnitude of the divergence between the two classes of basis coefficients. The divergence is evidently the largest in the mu-band. Figure 3 shows a comparison of classification results obtained using the proposed method to classification results obtained using the logarithm of the variances of the CSP (or CSSP) filtered signals as feature vectors. In both cases LIBLINEAR was used as classifier using leave-one-out cross-validation.

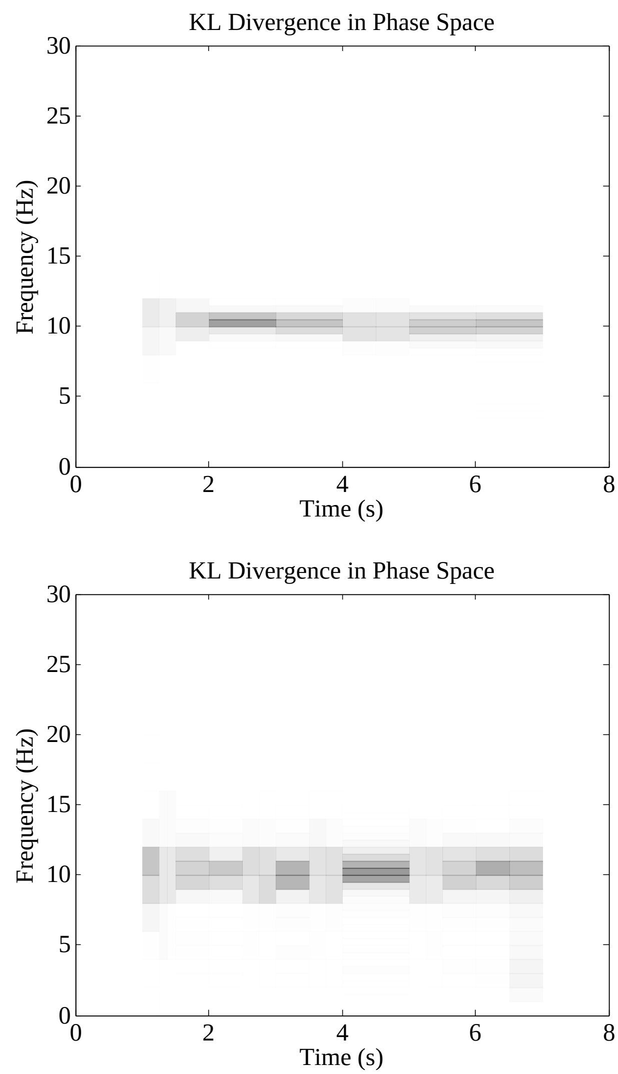

Fig. 2: Magnitude of the symmetrized Kullback-Leibler divergence in phase space for two of the subjects. The start of the visual cue corresponds to second 1 in the plots.

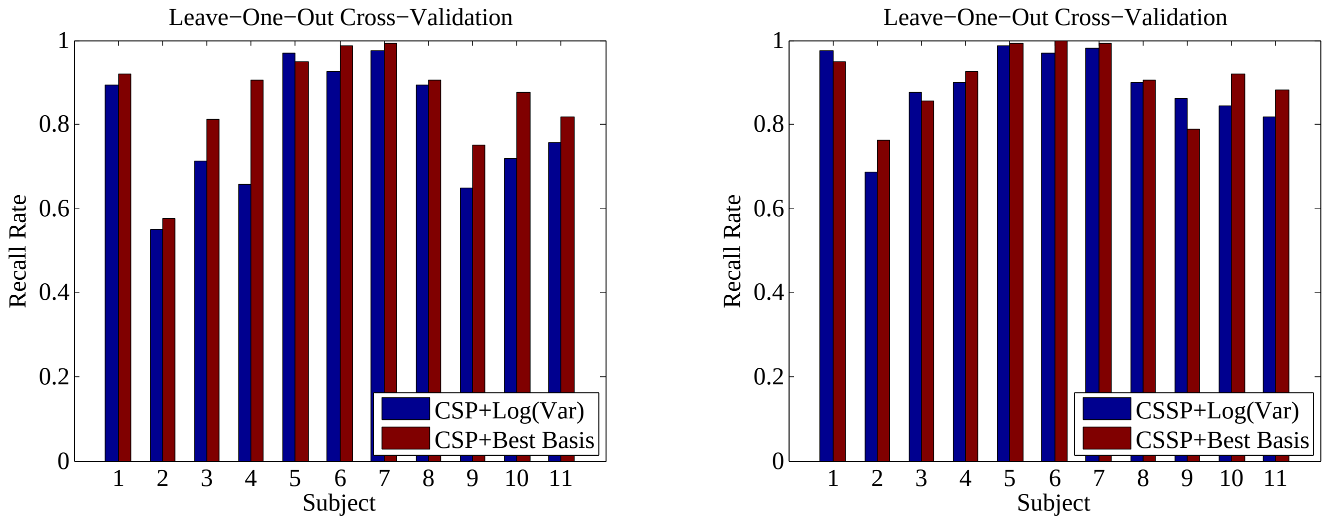

FIG. 3: Comparison of classification results using the output of the CSP filter (left) and using the output of the CSSP filter (right).

4 Conclusions

The presented method performs a pairwise comparison between signals belonging to two different classes to determine which time-frequency features maximize a specific divergence measure between the two classes. A test on motor imagery data shows that the method was able to detect differences in the mu-band but not in the beta-band. This may change if in the future we will be able to incorporate the spatial filtering into the *best basis* selection algorithm. As far as the classification accuracy is concerned, by means of the proposed method we were able to improve on the recall rates of 10 out of 11 subjects when using the CSP filtered data and on 8 out of 11 subjects when using CSSP. An interesting topic for further research will be the analysis of the robustness of the selected basis across different subjects. We would also like to investigate the impact of the size of the training set and the percentage of pairwise comparisons that are actually required to select a stable best basis.

Acknowledgments

This work was supported by the ANR project CO-ADAPT (ANR-09-EMER-002-05).

References

[1] R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. on Inf. Th.* **38**:2 (1992), pp. 713-718.

[2] J. Fruitet, M. Clerc and T. Papadopoulo. Preliminary study for an offline hybrid BCI using sensorimotor rhythms and beta rebound. *International Journal of Bioelectromagnetism*, **13**:2 (2011), pp. 70-71.

[3] S. Kullback and R. Leibler. On information and sufficiency. *Annals of Mathematical Statistics* **22**:1 (1951), pp. 79-86.

[4] L.F. Nicolas-Alonso and J. Gomez-Gilemail. Brain computer interfaces, a review. *Sensors*, **12**:2 (2012), pp. 1211-1279.

[5] N. Saito. *Local Feature Extraction and Its Applications Using a Library of Bases*, PhD thesis, Yale University (1994).

[6] D. Vautrin, X. Artusi, M-F. Lucas and D. Farina. A Novel Criterion of Wavelet Packet Best Basis Selection for Signal Classification with Application to Brain-Computer Interfaces, *IEEE Trans. on Biomed. Engineering* **56**:11 (2009-11), pp. 2734-2738.

[7] M.V. Wickerhauser. *Adapted Wavelet Analysis from Theory to Software*, A.K. Peters (1994).

[8] Z. J. Koles. The quantitative extraction and topographic mapping of the abnormal components in the clinical eeg. *Electroencephalography and Clinical Neurophysiology*, **79**:6 (1991), pp. 440-447.

[9] H. Ramoser, J. Muller-Gerking and G. Pfurtscheller. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. on Rehabilitation Eng.*, **8**:4 (2000), pp. 441-446.

[10] S. Lemm, B. Blankertz, G. Curio and K. Muller. Spatio-spectral filters for improving the classification of single trial EEG. *IEEE Trans. on Biomed. Engineering* **52**:9 (2005-9), pp. 1541-1548.

[11] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang and C.-J. Lin. LIBLINEAR: A library for large linear classification. *J. of Machine Learning Research* **9** (2008), pp. 1871-1874.